An Ensemble Adjustment Filter for Data Assimilation

Jeffrey L. Anderson
Geophysical Fluid Dynamics Laboratory
P.O. Box 308, Princeton, NJ 08542

Submitted to Monthly Weather Review
20 September, 2000
Revised 22 January, 2001

Abstract

A theory for estimating the probability distribution of the state of a model given a set of observations exists. This non-linear filtering theory unifies the data assimilation and ensemble generation problem that have been key foci of prediction and predictability research for numerical weather and ocean prediction applications. A novel Monte Carlo approximation to the non-linear filter is developed and applied in perfect model experiments in a low order model, and in a non-divergent barotropic model in both perfect and imperfect model applications. This ensemble adjustment method produces assimilations with small ensemble mean errors while providing reasonable measures of uncertainty in the assimilated variables. The method can assimilate observations with a nonlinear relation to model state variables. It can also use observations to estimate the value of imprecisely known model parameters. The method is shown to have significant advantages over four dimensional variational assimilation in low order models and scales easily to much larger applications. The key to the method is the way in which ensemble prior estimates are modified when observations become available. The ensemble adjustment method is compared to the ensemble Kalman filter which has been applied in atmospheric and oceanic science. While the methods are similar in many aspects, ensemble adjustment is able to produce better results in the cases examined. Noise introduced into the assimilated ensemble in the ensemble Kalman filter appears to be responsible for limiting its relative performance.

1. Introduction

Methods used to produce operational forecasts of the atmosphere have been undergoing a gradual evolution over the past decades. Prior to the 1980's, operational prediction centers attempted to produce a single 'deterministic' prediction of the atmosphere; initial conditions for the prediction were derived using an assimilation and initialization process that used, at best, information from a single earlier prediction. Since that time, the operational use of multiple forecasts, ensembles, has been developed in an attempt to produce information about the probability distribution (Van Leeuwen and Evensen1996) of the atmospheric forecast (Molteni et al 1996, Tracton and Kalnay 1993, Toth and Kalnay 1993, 1997, Houtekamer et al 1995).

Anderson and Anderson (1999, hereafter AA) developed a Monte Carlo implementation of the nonlinear filtering problem (Jazwinski 1970, chapter 6) for use in atmospheric data assimilation. The framework developed in AA allowed a synthesis of the data assimilation and ensemble generation problem. The method worked well in low-order systems, but it was not immediately clear how it could be applied to the vastly larger models that are commonplace for atmospheric and oceanic prediction and simulation.

The fundamental problem facing the AA method and a variety of other ensemble assimilation techniques, in particular the ensemble Kalman filter (Evensen 1994, Houtekamer and Mitchell 1998, Keppenne 2000), that have been proposed for atmospheric and ocean models is that the sample sizes of practical ensembles are far too small to give meaningful statistics about the complete distribution of the model state conditional on the available observations (Burgers et al 1998, van Leeuwen 1999). This has led to a variety of simplifications and heuristic methods that try to overcome this problem, for instance using ensembles to generate statistics for small subsets of the model variables (Evensen and Van Leeuwen 1996, Houtekamer and Mitchell 1998).

The AA method has a number of undesirable features when applied sequentially to small subsets of model state variables. The most pathological is that prior covariances between model state variables in different subsets are destroyed whenever observations are assimilated. A new method of updating the ensemble in a filter, called ensemble adjustment, is described here. This method retains many desirable features of the AA filter while allowing application to subsets of state variables. In addition, modifications to the filter design allow assimilation of observations that are related to the state variables by arbitrary non-linear operators as can be done with ensemble Kalman filters. The result is an ensemble assimilation method that can be applied efficiently to arbitrarily large models given certain caveats. Low-order model results to be presented here suggest that the quality of these assimilations is significantly better than those obtained by current state-of-the-art methods like four dimensional variational assimilation (Le Dimet and Talagrand 1986, Lorenc 1997, Rabier et al 1998) or ensemble Kalman filters. Although the discussion that follows is presented specifically in the context of atmospheric models, it is also applicable to other geophysical models like ocean or complete coupled climate system models.

2. An ensemble adjustment filter

A. Joint state / observation space non-linear filter

The state of the atmosphere, $\chi_{t,}$ at a time, t, has the conditional probability density function

$$\mathbf{p}(\chi_t \mid \mathbf{Y}_t), \tag{1}$$

where $\mathbf{Y_t}$ is the set of all observations of the atmosphere that are taken at or before time t. Following Jazwinski (1970) and AA, let $\mathbf{x}_\mathrm{t}$ be a discrete approximation of the atmospheric state which can be advanced in time using the atmospheric model equations:

$$d\mathbf{x}_\mathrm{t} / dt = M(\mathbf{x}_t, t) + \mathbf{G}(\mathbf{x}_t, t)\, \mathbf{w}_t. \tag{2}$$

Here, $\mathbf{x}_t$ is an n-dimensional vector that represents the state of the model system at time t, M is a deterministic forecast model, and $\mathbf{w}_t$ is a white Gaussian process of dimension r with mean 0 and covariance matrix $\mathbf{S}(t)$ while $\mathbf{G}$ is an n x r matrix. The second term on the right represents a stochastic component of the complete forecast model (2). In fact, all of the results that follow apply as long as the time update (2) is a Markov process. As in AA, the stochastic term is neglected initially. For most of this paper, the filter is applied in a perfect model context where

$$d\mathbf{x}_t / dt = M(\mathbf{x}_t, t) \tag{3}$$

exactly represents the evolution of the system of interest.

Assume that a set of m_t scalar observations, $\mathbf{y}^\mathrm{o}{}_t$, is taken at time t (the superscript ‘o’ stands for observations). The observations are functions of the model state variables and include some observational error (noise) which is assumed to be Gaussian (although the method can be extended to non-Gaussian observational error distributions):

$$\mathbf{y}^\mathrm{o}{}_t = \mathbf{h}_\mathrm{t}(\mathbf{x}_t, t) + \varepsilon_t(\mathbf{x}_\mathrm{t}, \mathrm{t}) \tag{4}$$

Here, $\mathbf{h}$ is an m_t-vector function of the model state and time and ε_t is an m_t-vector observational error selected from an observational error distribution with mean 0 and covariance R_t; m_t is the size of the observations vector which can itself vary with time. It is assumed that the ε_t for different times are uncorrelated. This may be a reasonable assumption for many traditional ground-based observations although other observations, for instance satellite radiances, may have significant temporal correlations in observational error.

The set of observations, $\mathbf{y}^\mathrm{o}{}_\mathrm{t}$, available at time t can be partitioned into the largest number of subsets, $\mathbf{y}^o_{t,k}$, for which the observational error covariance between subsets is negligible. Then,

$$\mathbf{y}^o_{t,k} = \mathbf{h}_{t,k}(\mathbf{x}_t, t) + \varepsilon_{t,k}(\mathbf{x}_t, t), \qquad k = 1, ..., r \tag{5}$$

where $\mathbf{y}^o_{t,k}$ is the kth subset at time t, $\mathbf{h}_{t,k}$ is an m-vector function (m can vary with both time and subset), $\varepsilon_{t,k}$ is an m-vector observational error selected from an observational error distribution with mean 0 and mxm covariance matrix $R_{t,k}$, and r is the number of subsets at time t. Many types of atmospheric observations have observational error distributions with no significant correlation to the error distributions of other contemporaneous observations leading to subsets of size one ($\mathbf{y}^o_{t,k}$ is scalar). Note that no restrictions have been placed on $\mathbf{h}_{t,k}$ (and $\mathbf{h}_t$); in particular the observed variables are not required to be linear functions of the state variables.

A cumulative observation set, $\mathbf{Y}_\tau$, can be defined as the superset of all observations, $\mathbf{y}^o_t$, for times $t <= \tau$. The conditional probability density of the model state at time *t*,

$$\mathbf{p}(\mathbf{x}_t \mid \mathbf{Y}_t), \tag{6}$$

is the complete solution to the filtering problem when adopting a Bayesian point of view (Jazwinski 1970). Following AA, the probability distribution (6) is referred to as the *analysis probability distribution* or *initial condition probability distribution*. The forecast model (3) allows the computation of the conditional probability density at any time after the most recent observation time:

$$\mathbf{p}(\mathbf{x}_t \mid \mathbf{Y}_\tau) \quad t>\tau. \tag{7}$$

This predicted conditional probability density is a forecast of the state of the model, and also provides the prior distribution at the time of the next available observations for the assimilation problem. The temporal evolution of this probability distribution is described by the Liouville equation as discussed in Ehrendorfer (1994). The probability distribution (7) will be referred to as the *first guess probability distribution* or *prior probability distribution* when used to assimilate additional data, or the *forecast probability distribution* when a forecast is being made.

$\mathbf{Y}_{\tau,\kappa}$ is defined as the superset of all observation subsets $\mathbf{y}^o_{t,k}$ with $t<=\tau$ and $k <= \kappa$, (note that $\mathbf{Y}_{t,0} = \mathbf{Y}_{t_p}$ where t_p is the previous time at which observations were available). Assume that the conditional probability distribution $p(\mathbf{x}_t \mid \mathbf{Y}_{t,k-1})$ is given. The conditional distribution after making use of the next subset of observations is:

$$\mathbf{p}(\mathbf{x}_t \mid \mathbf{Y}_{t,k}) = \mathbf{p}(\mathbf{x}_t \mid \mathbf{y}^o_{t,k}, \mathbf{Y}_{t,k-1}). \tag{8}$$

For k = 1, the forecast model (3) must be used to compute $\mathbf{p}(\mathbf{x}_t \mid \mathbf{Y}_{t_p})$ from $\mathbf{p}(\mathbf{x}_{t_p} | \mathbf{Y}_{t_p})$.

In preparation for applying the numerical methods outlined later in this section, define the joint state/observation vector (referred to as joint state vector) for a given t and k as $\mathbf{z}_{t,k} = \{\mathbf{x}_t, \mathbf{h}_{t,k}(\mathbf{x}, t)\}$, a vector of length n + m where m is the size of the observational subset $\mathbf{y}^o_{t,k}$. The idea of working in a joint state/observation space can be used in a very general description of the filtering problem (Tarantola 1987). Working in the joint space allows arbitrary observational operators, $\mathbf{h}$, to be used in conjunction with the ensemble methods developed below. Following the same steps that led to (8) gives

$$\mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{Y}_{t,k}) = \mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{y}^o_{t,k}, \mathbf{Y}_{t,k-1}). \tag{9}$$

Returning to the approach of Jazwinski, Bayes' rule gives

$$\mathbf{p}(\mathbf{z}_{t,k} | \mathbf{Y}_{t,k}) = \mathbf{p}(\mathbf{y}^o_{t,k} \mid \mathbf{z}_{t,k}, \mathbf{Y}_{t,k-1})\mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{Y}_{t,k-1}) / \mathbf{p}(\mathbf{y}^o_{t,k} \mid \mathbf{Y}_{t,k-1}). \tag{10}$$

Since the observational noise $\varepsilon_{t,k}$ is assumed uncorrelated for different observation times and subsets,

$$\mathbf{p}(\mathbf{y}^o_{t,k} \mid \mathbf{z}_{t,k}, \mathbf{Y}_{t,k-1}) = \mathbf{p}(\mathbf{y}^o_{t,k} \mid \mathbf{z}_{t,k}). \tag{11}$$

Incorporating (11) into (10) gives

$$\mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{Y}_{t,k}) = \mathbf{p}(\mathbf{y}^o_{t,k} \mid \mathbf{z}_{t,k})\, \mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{Y}_{t,k-1}) / \mathbf{p}(\mathbf{y}^o_{t,k} \mid \mathbf{Y}_{t,k-1}) \tag{12}$$

which expresses how new sets of observations modify the prior joint state conditional probability distribution available from predictions based on previous observation sets. The denominator is a normalization that guarantees that the total probability of all possible states is 1. The numerator is a product of two terms, the first representing new information from observation subset k at time *t* and the second representing the prior constraints. The prior term gives the probability that a given model joint state, $\mathbf{z}_{t,k}$, occurs at time *t* given information from all observations at previous times and the first k-1 observation subsets at time t. The first term in the numerator of (12) evaluates how likely it is that the observation subset $\mathbf{y}^o_{t,k}$ would be taken given that the state was $\mathbf{z}_{t,k}$. This algorithm can be repeated recursively until the time of the latest observation, at which point (3) can be used to produce the forecast probability distribution at any time in the future.

B. Computing the filter product

Applying (12) to large atmospheric models leads to a number of practical constraints. The only known computationally practical way to advance the prior state distribution, $\mathbf{x}_t$, in time is to use Monte Carlo techniques (ensembles). Each element of a set of states sampled from (6) is advanced in time independently using the model (3). The observational error distributions of most climate system observations are poorly known and are generally given as Gaussians (i.e. a standard deviation or covariance).

Assuming that (12) must be computed given an ensemble sample of $\mathbf{p}(\mathbf{x}_t \mid \mathbf{Y}_{t,k\text{-}1})$, an ensemble of the joint state prior distribution, $\mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{Y}_{t,k\text{-}1})$, can be computed by applying $\mathbf{h}_{t,k}$ to each ensemble sample of $\mathbf{x}_t$. The result of (12) must be an ensemble sample of $\mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{Y}_{t,k})$. As noted in AA, there is generally no need to compute the denominator (the normalization term) of (12) in ensemble applications. Four methods for approximating the product in the numerator of (12) are presented, all using the fact that the product of two Gaussian distributions is itself Gaussian and can be computed in a straightforward fashion.

i. Gaussian ensemble filter

This is an extension of the first filtering method described in AA to the joint state space. Let $\bar{z}^p$ and Σ^p be the sample mean and covariance of the prior joint state, $\mathbf{p}(\mathbf{z}_{t,k} \mid \mathbf{Y}_{t,k\text{-}1})$, ensemble. The observation subset $\mathbf{y}^o = \mathbf{y}^o_{t,k}$ has error covariance $\mathbf{R} = \mathbf{R}_{\mathbf{t,k}}$ ($\mathbf{R}$ and $\mathbf{y}^o$ are functions of the observational system). The expected value of the observation subset given the state variables is $\mathbf{h}_{t,k}(\mathbf{x}_t, t)$ as in equation (5), but in the joint state space this is equivalent to the simple m x n+m linear operator $\mathbf{H}$, where $H_{k,k+n} = 1.0$ for k=1, ..., m and all other elements of $\mathbf{H}$ are 0, so that the estimated observation values calculated from the joint state vector are $\mathbf{y}_{t,k} = \mathbf{H}\mathbf{z}_{t,k}$.

Assuming that the prior distribution can be represented by a Gaussian with the sample mean and variance results in the numerator of (12) having covariance

$$\Sigma^u = [(\Sigma^p)^{-1} + H^T \boldsymbol{R}^{-1} H]^{-1}, \tag{13}$$

mean

$$\overline{z^u} = \Sigma^u[(\Sigma^p)^{-1}\overline{\boldsymbol{z}^p} + H^T\boldsymbol{R}^{-1}\boldsymbol{y}^o], \tag{14}$$

and a weight

$$\boldsymbol{D} = (\overline{\boldsymbol{z}^p})^T(\Sigma^p)^{-1}\overline{\boldsymbol{z}^p} + (\boldsymbol{y}^o)^T\boldsymbol{R}^{-1}\boldsymbol{y}^o - (\overline{\boldsymbol{z}^u})^T(\Sigma^u)^{-1}\overline{\boldsymbol{z}^u}. \tag{15}$$

These are an extension of eqs. A1-A4 in AA to the joint state space (S. Anderson, personal communication). In the Gaussian ensemble filter, the updated ensemble is computed using a random number generator to produce a random sample from a Gaussian with the covariance and mean from (13) and (14). The *expected values* of the mean and covariance of the resulting ensemble are $\overline{z}^u$ and Σ^u while the expected values of all higher order moments should be 0.

ii. Kernel filter

The kernel filter mechanism developed in AA can also be extended to the joint state space. In this case, the prior distribution is approximated as the sum of N Gaussians with means z^p_i and identical covariances Σ^p, where z^p_i is the ith ensemble sample of the prior and N is the ensemble size. The product of each Gaussian with the observational distribution is computed by applying equation (13) once and equations (14-15) N times, with $\overline{z}^p$ replaced by z^p_i in (14) and (15) and $\overline{z}_u$ being replaced by z^u_i in (15) where z^u_i is the result of the ith evaluation of (14). The result is N new distributions with the same covariance but different means and associated weights whose sum represents the product. An updated ensemble is generated by randomly sampling from this set of distributions as in AA. In almost all cases, the *values* and *expected values* of the mean and covariance and higher order moments of the resulting ensemble are functions of higher order moments of the prior distribution. This makes the kernel filter potentially more general than the other three methods, however, computational efficiency issues outlined later appear to make it impractical for application in large models.

iii. Ensemble Kalman Filter (EKF)

The ensemble Kalman filter (EKF hereafter) forms a random sample of the observational distribution, $\mathbf{p}(\mathbf{y}^{o}_{t,k} \mid \mathbf{z}_{t,k})$ in (12), sometimes referred to as ‘perturbed observations’ (Houtekamer and Mitchell 1998) using a random number generator to sample the observational error distribution, $\mathbf{v}_{t,k}(\mathbf{x}_t, t)$, and adding these samples to the observation, $\mathbf{y}^o$, to form an ensemble sample of the observation distribution, y_i, i = 1, ... ,N. Equation (13) is computed once to find the value of Σ^u. Equation (14) is evaluated N times to compute z^u_i, with $\bar{\mathbf{z}}^p$ and $\mathbf{y}^o$ replaced by the z^p_i and y_i, where the subscript refers to the value of the ith ensemble member. As shown in Burgers et al (1998), computing a random sample of the product as the product of random samples is a valid Monte Carlo approximation to the nonlinear filtering equation (12). Essentially, the EKF can be regarded as an ensemble of Kalman filters, each using a different sample estimate of the prior mean and observation. The *expected values* of the sample mean and covariance of the updated ensemble are $\bar{z}^u$ and Σ^u, but the expected values of higher order moments are functions of higher order moments of the prior distribution.

Deriving the EKF directly from the nonlinear filtering equation (12) may be more transparent than some derivations found in the EKF literature where the derivation begins from the statistically linearized Kalman filter equations. This traditional derivation masks the statistically nonlinear capabilities of the EKF, for instance, the fact that both prior and updated ensembles can have an arbitrary (non-Gaussian) structure. Additional enhancements to the EKF, for instance the use of two independent ensemble sets (Houtekamer and Mitchell 1998), can also be developed in this context.

iv. Ensemble Adjustment Filter (EAF)

In the new method that is the central focus of this paper, equations (13-14) are used to compute the mean and covariance of the updated ensemble. A new ensemble that has exactly these sample characteristics while maintaining as much as possible the higher moment structure of the prior distribution is generated directly. The method, referred to as *ensemble adjustment*, for generating the new ensemble applies a linear operator, **A**, to the prior ensemble in order to get the updated ensemble.

$$\boldsymbol{z}_i^u = \boldsymbol{A}^T(\boldsymbol{z}_i^p - \overline{z^p}) + \overline{\boldsymbol{z}^u}, \text{ i} = 1, ..., \text{N} \tag{16}$$

where $\boldsymbol{z}_i^p$ and z_i^u are individual members of the prior and updated ensemble. **A** is selected so that the sample covariance of the updated ensemble is identical to that computed by (13). Appendix A demonstrates that **A** exists (many **A**'s exist since corresponding indices of prior and updated ensemble members can be scrambled) and discusses a method for computing the appropriate **A**.

Since there are a number of approximations permeating the ensemble adjustment algorithm, there are naturally inaccuracies in the prior sample covariance and mean. As for other filter implementations, like the Kalman filter, sampling error or other approximations can cause the computed prior covariances to be too small at some times. The result is that less weight is given to new observations when they become available resulting in further reduced covariance in the next prior estimate. Eventually, the prior may no longer be impacted significantly by the observations, and the assimilation will depart from the observations. A number of sophisticated methods for dealing with this problem can be developed. Here, only a simple remedy is used. The prior covariance matrix is multiplied by a constant factor, usually slightly larger than one. If there are some local (in phase space) linear balances between the state variables on the model's attractor, then the application of small covariance inflation might be expected to maintain these balances while still increasing uncertainty in the state estimate. Clearly, even if there are locally linear balanced aspects to the dynamics on the attractor, the application of sufficiently large covariance inflations would lead to significantly unbalanced ensemble members.

The *covariance inflation* factor is selected empirically here in order to give a filtering solution that does not diverge from the observations while keeping the prior covariances small. More sophisticated approaches to this problem are necessary when dealing with models that have significant systematic errors (i.e. when assimilating real observations) and are currently being developed.

An alternate approach to deter filter divergence is to add a random (Gaussian) noise in phase space, but this would clearly lead to large projections off any locally linear attractor. The EKF generally does not suffer from filter divergence, because noise is being added directly through the use of perturbed observations. However, this noise may also destroy information in the prior distribution as discussed in later sections.

C. Simplifications to reduce computational cost

The size of atmospheric models and of computationally affordable ensembles necessitate additional simplifications when computing updated means and covariances. The sample prior covariance computed from an N-member ensemble is non-degenerate in only N - 1 dimensions of the joint state space. If the global covariance structure of the assimilated joint state cannot be represented accurately in a subspace of size N-1, filter methods are unlikely to work without making use of other information about the covariance structure (Miller et al 1994a). When the perfect model assumption is relaxed, this can become an even more difficult problem since model systematic error is not necessarily likely to project on the subspace spanned by small ensembles.

One approach to dealing with this degeneracy is to project the model state onto some vastly reduced subspace before computing products, leading to methods like a variety of reduced space (ensemble) Kalman filters (Kaplan et al 1997, Gordeau et al 2000, Brasseur et al 1999). A second approach, used here, is to update small sets of 'physically close' state variables independently. Associated with each of these sets of state variables is a number of 'possibly related' observations (using only 'possibly related' observations when updating state variables is similar to the cutoff radius for observations suggested by Houtekamer and Mitchell (1998) and itself reduces computational cost) and a second set of additional related state variables.

Let C be a set containing the indices of all state variables in a particular independent subset of state variables, referred to as a *compute domain*, along with the indices of all possibly related observations in the current joint state vector. Let D be a set containing the indices of all additional related state variables, referred to as the *data domain*. Then $\Sigma^u_{i,j}$ and $\overline{z^u_i}$ where $i, j \in C$ are computed using an approximation to $\Sigma^p_{i,j}$ in which all terms for which $i \notin C \cup D$ or $j \notin C \cup D$ are set to zero. In other words, the state variables in each compute domain are updated making direct use only of prior covariances between themselves and also variables in the corresponding data domain and the related state variables. These subsets can be computed statically (as will be done in all applications here) or dynamically using information available in the prior covariance and possibly additional *a priori* information. The data domain state variables in

D may themselves be related strongly to other state variables outside of $C \cup D$ and so are more appropriately updated in conjunction with some other compute set.

Additional computational savings can accrue by performing a singular value decomposition on the prior covariance matrix (already done as part of the numerical method for updating the ensembles as outlined in Appendix A) and working in a subspace spanned by singular vectors with non-negligible singular values. This singular vector filtering is a prerequisite if the size of the set $C \cup D$ exceeds N-1, leading to a degenerate sample prior covariance matrix (Houtekamer and Mitchell 1998, Evensen and van Leeuwen 1996).

All results in the following use particularly simple and computationally efficient versions of the filtering algorithms. First, all observation subsets contain a single observation; in this perfect model case this is consistent with the observations which have zero error covariance with other observations. Second, the compute domain set, *C*, also contains only a single element and the data domain D is the null set in all cases. The result is that each component of the mean and each prior covariance diagonal element is updated independently (this does NOT imply that the prior or updated covariances are diagonal). The joint state prior covariance matrix used in each update is 2x2 containing the covariance of a single state and the single observation in the current observational subset. In computing the products to get the new state estimate, the EAF algorithm used here only makes use of singular value decompositions and inverses of 2x2 matrices (the three other filter methods similarly require only 2x2 matrix operations). Allowing larger compute and data domains would generally be expected to improve the results discussed in later sections while leading to significantly increased constant factors multiplying computational cost.

D. Motivation for ensemble adjustment algorithm

This subsection discusses advantages of the EAF over the Gaussian and kernel filters, both referred to as *resampling* Monte Carlo (or just resampling) methods since a random sample of the updated distribution must be formed at each update step. Applying resampling filters locally to subsets of the model state variables as discussed in the previous subsection, one might expect the structure of the assimilated probability distributions to be simpler and more readily approximated by Gaussians. Subsets of state variables of size smaller than N can be used so that the problem of degenerate sample covariance matrices is avoided altogether. This can solve problems of filter divergence that result from global applications of resampling filters (AA). The state variables can be partitioned into compute and data subsets as described above, motivated by the concept that most state variables are closely related only to a subset of other state variables, usually those that are physically nearby. Ignoring prior covariances with more remote variables is expected to have a limited impact on the computation of the product. Similar approaches have been used routinely in EKFs (Houtekamer and Mitchell 1998).

Unfortunately, resampling ensemble filters are not well-suited for local application to subsets of state variables. Whenever an observation is incorporated, the updated mean(s) and covariance(s) are computed using eqs. (13) and (14) and a new ensemble is formed by randomly sampling the result. Even when observations with a very low relative information content (very large error covariance compared to the prior covariance) are assimilated, this resampling is done. However, resampling destroys all information about prior covariances between state variables in different compute subsets. The assumption that the prior covariances between different subsets are small is far from rigorous in applications of interest, so it is inconvenient to lose all of this information every time observations become available.

Figure 1a shows an idealized representation of a system with state variables X_1 and X_2 that are in different compute domains. An idealized observation of X_1 with Gaussian error distribution is indicated schematically by the density plot along the X_1 axis. Figure 1d shows the result of applying an EAF in this case. The adjustment pulls the value of X_1 for all ensemble members toward the observed value. The covariance structure between X_1 and X_2 is mostly preserved as the values of X_2 are similarly pulled inward. The result is qualitatively the same as applying a filter to X_1 and X_2 simultaneously (no subsets). Figure 1c shows the results of applying a single Gaussian resampling filter and Fig. 1b the result of a multiple kernel resampling filter as in AA. The resampling filters destroy all prior information about the covariance of X_1 and X_2.

There are other related problems with resampling ensemble filters. First, it is impossible to meaningfully trace individual assimilated ensemble trajectories in time. While the EAF maintains the relative positions of the prior samples, the letters in Figs. 1b and 1c are scrambled throughout the resulting distributions. This can complicate diagnostic understanding of the assimilation.

Second, if only a single Gaussian kernel is being used to compute the product, all information about higher order moments of the prior distribution is destroyed each time data is assimilated (Fig. 1c). AA introduced the sum of Gaussian kernels approximation to avoid this problem. In Fig. 1b, the projection of higher order structure on the individual state variable axes is similar to that in Fig. 1d, but the distribution itself winds up being qualitatively a quadrupole because of the loss of covariance information between X_1 and X_2.

These deficiencies of the resampling ensemble filters occur because a random sampling of the updated probability distribution is used to generate the updated ensemble. In contrast, the EAF retains some information about prior covariances between state variables in separate compute subsets as shown schematically in Fig. 1d. For instance, observations that have a relatively small information content make small changes to the prior distributions. Most of the covariance information between variables in different subsets survives the product step in this case. This is particularly relevant since the frequency of atmospheric and oceanic observations for problems of interest may lead to individual (subsets of) observations making relatively small adjustments to the prior distributions.

Ensemble adjustment also preserves information about higher order moments of prior probability distributions as shown in Fig. 1d. Again, this information is particularly valuable when observations make relatively small adjustments to the prior distributions. For instance, if the dynamics of a model are generating distributions with interesting higher moment structure, for instance a bimodality, this information can survive the update step using the EAF but is destroyed by resampling with a single Gaussian kernel.

Individual ensemble trajectories can be meaningfully traced through time with the EAF (see also Figs. 3, 6 and 7). If observations make small adjustments to the prior, individual ensemble members look similar to free runs of the model with periodic small jumps where data is incorporated. Note that the adjustment method is deterministic after initialization, requiring no generation of random numbers once an initial ensemble is created.

The EAF is able to eliminate many of the shortcomings of the resampling filters. Unlike the resampling filters, it can be applied when subsets of state variables are used for computing updates. Ensemble adjustment retains information about higher order moments of prior distributions and individual ensemble trajectories are more physically relevant leading to easier diagnostic evaluation of assimilations. All of these advantages are particularly pronounced in instances where observations at any particular time have a relatively small impact on the prior distribution, a situation that seems to be the case for most climate system model / data problems of interest.

E. Advantages over Ensemble Kalman Filter

The EKF shares most of the advantages of the EAF as discussed in the previous subsections. However, as demonstrated in later sections, the EKF does not perform as well as the EAF. The primary reason is that the random sampling of the observational distribution (perturbed observations) used in the EKF introduces a small but significant noise into the product. This noise acts to destroy the prior relations between state variables in different compute domains, although to a much lesser degree than with resampling methods. Nevertheless, the persistent introduction of this noise is, at least in the cases examined, sufficient to degrade the relative performance of the EKF. This degradation might be expected to be particularly serious in models that have a so-called ‘slow manifold’, where the behavior of on and off-attractor dynamics is significantly different. Investigation of the relative performance of the EAF and EKF in primitive equation models which have behavior of this type should be able to give more insight into how serious the introduction of noise by the EKF would be in this case.

3. Results from a low order system

The EAF is applied to the 40-variable model of Lorenz (1996) (Appendix B; referred to hereafter as L96) which was used for simple tests of targeted observation methodologies in Lorenz and Emanuel (1998). The number of state variables is greater than the smallest ensemble sizes (approximately 10) required for usable sample statistics and the model has a number of physical characteristics similar to those of the real atmosphere.

Synthetic observations of a long control run of the L96 model are used. In the first case examined, the observational operator, **h**, is the identity (each state variable is observed directly), the observational error covariance is diagonal with all elements 4.0 (observations have independent error variance of 4), and observations are available every timestep. A 20 member ensemble is used and results are analyzed in detail for assimilation steps 101 to 200. The covariance inflation factor is set to 1.05 and all observations are allowed to impact each state variable.

As discussed in detail in AA, the goal of filtering is to produce an ensemble with small ensemble mean error and with the true state being statistically indistinguishable from a randomly selected member of the ensemble. The time mean RMS error of the ensemble mean for this assimilation is 2.57, a significant improvement over the expected error of $12.65 = (40*4)^{0.5}$ that would result from just using the raw observations as an estimate of the state at each time. Figure 2 shows the RMS error of the ensemble mean for this assimilation and for forecasts started from the assimilation out to leads of 20 assimilation times.

Figure 3a shows a time series of the ‘truth’ from the control run and the corresponding ensemble members (ten of the total of twenty are displayed to reduce clutter) and ensemble mean from the EAF for variable X_1. There is no evidence in this figure that the assimilation is inconsistent with the truth. The truth lies close to the ensemble mean (compared to the range of the variation in time) and generally is inside the 10 ensemble members plotted. The ensemble spread varies significantly in time; for instance, the ensemble is more confident about the state (less spread) when the wave peak is approaching at assimilation time 133 than when the trough passes at time 123. The ability to trace individual ensemble member trajectories in time is also clearly demonstrated; as noted in section 2 this could not be done in resampling methods. As an example, notice the trajectory that maintains a consistently high estimate from steps 118 through steps 127.

Figure 4 displays the RMS error of the ensemble mean and the ensemble spread (the mean RMS difference between ensemble members and the ensemble mean) for the X_1 variable for assimilation times 100 to 200. The time variability of the RMS error of the ensemble mean and the ensemble spread have a correlation of 0.62 demonstrating a significant relation (at the 95% confidence level) between ensemble spread and ensemble mean error in this assimilation. The expected relation between spread and skill (Murphy 1988, Barker 1991) will be analyzed in detail in a follow-on study.

Figure 5 shows the result of forming a rank histogram (a Talagrand diagram, Anderson 1996) for the X_1 variable over the same period shown in Fig. 4. At each analysis time, this technique uses the order statistics of the analysis ensemble of a scalar quantity to partition the real line into n+1 intervals (bins); the truth at the corresponding time falls into one of these n+1 bins. A necessary condition for the analysis ensemble to be a random sample of (6) is that the distribution of the truth into the n+1 bins be uniform (Anderson 1996). This is evaluated with a standard chi-square test applied to the distribution of the truth in the n+1 bins. The null hypothesis is that the truth and analysis ensemble are drawn from the same distribution. Fig. 5 does not show much of the pathological behavior demonstrated by inconsistent ensembles, for instance clumping in a few central bins or on one or both wings. The chi-square test applied to Fig. 5 gives 25.62, indicating an 82 percent chance that the ensemble was selected from a different distribution than the truth for this sample of 100 assimilation times. Obviously, if one uses large enough samples the truth will always be significantly different from the ensemble at arbitrary levels of confidence. In addition, the bins occupied by the truth on successive timesteps are not independent (see for instance Fig. 3a) as is assumed by the chi-square test. This can lead to chi-square results that assume too many degrees of freedom and indicate that the distribution is less uniform than it is in reality (T. Hamill, personal communication).

Another simple method for evaluating whether the truth is similar to a randomly selected ensemble member is to compute the ratio of the time-averaged RMS error of the ensemble mean to the time-averaged mean RMS error of the ensemble members (this can be done for the complete state vector or individual state components). As shown by Murphy (1988, 1990), this ratio (referred to as the RMS ratio hereafter) should be

$$R = \sqrt{(N+1)/2N}$$

if the truth is statistically indistinguishable from a member of the analysis ensemble for an N-member ensemble. For this assimilation, the ratio of R for the complete state vector to the expected value of R is 0.99, close to unity but indicating that the ensemble has too much uncertainty (slightly too much spread).

The same experiment has been run using only a 10-member ensemble. Results are generally slightly worse, as shown by the RMS error curves as a function of lead time in Fig. 2. Using ensembles smaller than 10 leads to sample covariance estimates that are too poor for the filter to converge. Using ensembles larger than 20 leads to small improvements in the RMS errors (not shown).

It is important to examine the rate at which ensemble mean error and spread grow if the assimilation is turned off to verify that the EAF is performing in a useful fashion. In this case, the forecast error growth plot in Fig. 2 shows that error doubles in about 12 assimilation times.

For comparison, the EKF is applied to the same observations from the L96 model and produces an assimilation with a time mean RMS error of 3.41, about 33 percent larger than for the EAF. Time series of the individual assimilation members for the EKF (Fig. 3b) are considerably noisier than those for the EAF and in some places (like the one shown in Fig. 3b) it becomes difficult to meaningfully trace individual ensemble members in time. Apparently the EKF's addition of random noise through 'perturbed observations' at each assimilation step is sufficient to degrade the quality of the assimilation. The L96 system is quite tolerant of added noise with off-attractor perturbations decaying relatively quickly and nearly uniformly toward the attractor; it is likely that the noise added in the EKF will be of additional concern in less tolerant systems. An EKF with two independent ensemble sets (Houtekamer and Mitchell 1998) was also applied to this example. Results for a pair of 10 member EKF ensembles were worse than for the single 20 member ensemble.

A second test in the L96 model appraises the EAF's ability to deal with nonlinear forward observation operators. Forty observations placed randomly in the model domain are taken at each assimilation time. The observational operator, **h**, involves a linear interpolation from the model grid to the location of the observation, followed by a squaring of the interpolated value. The observational errors are independent with variance 64.0. The statistics of this experiment are somewhat more variable in time than those from the experiment described above, so results are presented for 2000 assimilation steps starting after a 200 step spin-up period. In this case, the EAF with covariance inflate of 1.12 produces a time mean RMS error of 2.90 while the ratio of the RMS ratio to its expected value is 0.96 indicating that the ensemble has somewhat too much spread. The results of the EAF in this case are qualitatively similar to those discussed in a related assimilation experiment which is discussed in more detail in the next section.

The EKF was also applied in this nonlinear observations case. The EKF assimilation rapidly diverged from the 'truth' and produced ensembles that looked similar to sets of trajectories randomly selected from the L96 model's climatology. Further investigation revealed that one problem was that the EKF was having difficulty using information from observations that were not significantly correlated with the state variable being updated, consistent with discussion in Houtekamer and Mitchell (1998). If observations were only allowed to impact state variables within a given distance, the behavior of the EKF was improved. Defining the width of the L96 domain as 1.0 (state variables are then separated by a distance of 0.025), allowing observations to impact only state variables within a distance of 0.11 gave the best results for the EKF. In this case, the EKF assimilation produced a time mean RMS error of 3.47 over assimilation steps 200 to 1200, however, even in this case the EKF diverged from the 'truth' shortly after assimilation step 1200. While additional heuristic modifications might be introduced to improve the EKF performance, the fundamental problem of noise introduced by 'perturbed' observations appears to limit its capabilities relative to the EAF in this model.

Although the EAF assimilation does not improve if the radius of impact of observations is reduced in this example, this should not be viewed too generally. If a large enough set of remote observations is allowed to impact each state variable the EAF assimilation quality will eventually degrade as spurious correlations with remote observations lead to erroneous impacts on state variables. The 40-variable model is not big enough to allow this behavior to be seen in the EAF. In large models, some method to limit the remote impact of observations is an essential part of both the EAF and EKF.

4. Estimation of model parameters

Most atmospheric models have many parameters (in dynamics and sub-grid scale physical parameterizations) for which appropriate values are not known precisely. One can recast these parameters as independent model variables (Derber 1989), and use assimilation to estimate values for the unknown parameters. The EAF produces a sample of the probability distribution of such parameters given available observations.

To demonstrate this capability, the forcing parameter, F, in the L96 model is treated as a model variable (the result is a 41 variable model) and the EAF is applied to the extended model using the same set of observations as in the non-linear observation case described in the last paragraph of section 3. For assimilation steps 200 to 2200, the EAF with covariance inflate of 1.10 produces a time mean RMS error of 2.69 while the ratio of the RMS ratio to its expected value is 0.96 indicating that the ensemble has somewhat too much spread. There is no good benchmark available to which these values can be compared, but they suggest that the EAF is working appropriately in this application. It is interesting to note that the RMS error is less than was obtained in the experiment at the end of section 3 in which F was fixed at the correct value. It seems possible that the introduction of additional uncertainty in this controlled fashion helps the EAF to function better.

Figure 6a shows a time series of F from this assimilation. The 'true' value is always 8, but the filter has no *a priori* information about the value or that the value is constant in time. Also, there are no observations of F, so information is available only indirectly through the nonlinear observations of the state variables; all observations are allowed to impact F. The ensemble mean assimilated value of F varies from about 7.8 to 8.3 during this segment of the assimilation; the time mean error of the ensemble mean of F over this period is 0.07. The assimilation is more confident about the value of F at certain times like time 795 than at others like time 775. The chi-square for F over the interval from assimilation times 700 to 800 (partly shown in Fig. 6) is very large indicating that the truth was selected from a different distribution. However, as shown in Fig. 6a there is a very large temporal correlation in which bin is occupied by the 'truth', suggesting that the number of degrees of freedom in the chi-square test would need to be modified to produce valid confidence estimates. Figure 7 shows a time series of this assimilation for variable X_1 from assimilation times 750 to 800. The assimilation tracks the truth with ensemble spread varying significantly with time, from relatively large uncertainty near time 760 to small uncertainty around time 780. The RMS error to ensemble spread correlation for X_1 is 0.283.

Estimating state variables in this way may offer a mechanism for tuning parameters in large models (Houtekamer and Lefaivre 1997, Mitchell and Houtekamer 2000), or even allow them to be time varying with a distribution. It remains an open question whether there is sufficient information in available observations to allow this approach in current generation operational models. Given the extreme difficulty of tuning sets of model parameters, an investigation of the possibility that this mechanism could be used seems to be of great importance.

One could further extend this approach by allowing a weighted combination of different sub-grid scale parameterizations in each ensemble member and assimilating the weights in an attempt to determine the most appropriate parameterizations. This would be similar in spirit to the approaches described by Houtekamer et al (1996) and might be competitive with methods of generating 'super-ensembles' from independent models (Goerss 2000, Harrison et al 1999, Evans et al 2000, Krishnamurthi et al 1999, Ziehmann 2000, Richardson 2000).

The best EKF result for this problem limited the impact of observations to state variables within a distance of 0.06 and had an RMS error of 3.95 over steps 200-2200. The EKF diverged for values of the observation impact distance cut-off that were much bigger or smaller than 0.06. This sensitivity of the EKF to the radius of impact of observations and the relative quality of the EAF again suggest that the noise introduced in the EKF is having an adverse impact on the assimilation. Figure 6b shows a time series of the EKF estimation of the forcing variable, F, for comparison with Fig. 6a. The individual EKF trajectories display a much greater high frequency time variation than did those for the EAF. Again, it can become difficult to consider tracing individual trajectories in time in a meaningful fashion. Consistent with the EAF, the EKF does work better when F is treated as an unknown parameter in this case than when it was fixed at its ‘true’ value of 8.0.

5. Comparison to Four Dimensional Variational Assimilation

Four dimensional variational assimilation methods (4D-var) are generally regarded as the present state-of-the-art for the atmosphere and ocean (Tziperman and Sirkes 1997). A 4D-var has been applied to the L96 model and results compared to those for the EAF. The 4D-var uses the L96 model as a strong constraint (Zupanski 1997), perhaps not much of an issue in a perfect model assimilation. The 4D-var optimization is performed with an explicit finite-difference computation of the derivative, with 128-bit floating point arithmetic, and uses as many iterations of a pre-conditioned, limited memory quasi-Newton conjugate gradient algorithm (NAG subroutine E04DGF) as are required to converge to machine precision (in practice the number of iterations is generally less than 200). The observations available to the 4D-var are identical to those used by the EAF, and the number of observation times being fit by the 4D-var is varied from 2 to 15 (cases above 15 began to present problems for the optimization even with 128-bits).

Figure 2 compares the RMS error of the 4D-var assimilations and forecasts to the those for the EAF assimilations out to leads of 20 assimilation times for the first case presented in section 3. All results are the mean for 101 separate assimilations and subsequent forecasts, between assimilation steps 100 to 200. As the number of observation times used in the 4D-var is increased, error is reduced but always remains much greater than the EAF error. The 4D-var cases also show accelerated error growth as a function of forecast lead compared to the EAF when the number of observation times for the 4D-var gets large, a symptom of increasing overfitting of the observations (Swanson et al 1998). An EAF with only 10 ensemble members is still able to outperform all of the 4D-var assimilations (Fig. 2).

The EAF outperforms 4D-var by using more complete information about the distribution of the prior. In addition to providing better estimates of the state, the EAF also provides information about the uncertainty in this estimate through the ensemble as discussed in section 3. Note that recent work by Hansen and Smith (2000) suggests that combining the capabilities of 4D-var and ensemble filters may lead to a hybrid that is superior to either.

6. Ease of implementation and performance

Implementing the EAF (or the EKF) requires little in addition to a forecast model and a description of the observing system. The implementation of the filtering code described here makes use of only a few hundred lines of Fortran-90 in addition to library subroutines to compute standard matrix and statistical operations. There is no need to produce a linear tangent or adjoint model (a complicated task for large models, Courtier et al 1993) nor are any of the problems involved with the definition of linear tangents in the presence of discontinuous physics an issue (Vukicevic and Errico 1993, Miller et al 1994b) as they are for 4D-var methods.

The computational cost of the filters has two parts: production of an ensemble of model integrations; computation of the filter products. Integrating the ensemble multiplies the cost of the single model integration used in some simple data assimilation schemes by a factor of N. In many operational atmospheric modeling settings, ensembles are already being integrated with more conventional assimilation methods so there may be no incremental cost for model integrations.

As implemented here, the cost of computing the filter products at one observation time is O(mnN) where m is the number of observations, n is the size of the model, and N is the ensemble size. The impact of each observation on each model variable is evaluated separately here. The computation for a given observation and state variable requires computing the 2x2 sample covariance matrix of the state variable and the prior ensemble observation, an O(N) operation repeated O(mn) times. In addition, several matrix inverses and singular value decompositions for 2x2 matrices are required (cost is not a function of m, n, or N). The computation of the prior ensembles of observed variables for the joint state/observation vector is also required, at a cost of O(m). It is difficult to envision an *ensemble* scheme that has a more favorable computational scaling than the O(mnN) for the methods applied here.

The constant factors in the cost can be reduced by making additional assumptions about the influence of observations. For instance, observations can be limited to impacting only state variables that are within a given physical distance of the observation as was required for the EKF in sections 3 and 4. In addition, run-time checks on the correlation between a state variable and a prior observed ensemble variable can be applied and if the correlation is small, the product for this pair need not be performed. Similar attempts to reduce cost (or improve performance) by limiting observation impact are common in EKF applications (Houtekamer and Mitchell 1998) and in other assimilation techniques.

7. Filter assimilation in barotropic model

The limitations of the resampling filter in AA made it impossible to apply to large systems with reasonable ensemble sizes. In this section, initial application of the EAF to a larger model are described. The model is a barotropic vorticity equation on the sphere, represented as spherical harmonics with a T42 truncation (Appendix C). The assimilation uses the streamfunction in physical space on a 64 latitude by 128 longitude grid (total of 8192 state variables).

The first case examined is a perfect model assimilation in which a long control run of the T42 model is used as the ‘truth’. To maintain variability, the model is forced as noted in the appendix. Observations of streamfunction are available every 12 hours at 250 randomly chosen locations on the surface of the sphere excluding the longitude belt between 60E and 160E where there are no observations. An observational error with standard deviation 1×10^6 $m^2 s^{-1}$ is added independently to each observation. A covariance inflation factor of 1.01 is used with a 20 member ensemble. In addition, only observations within 10 degrees of latitude and $\cos^{-1}(\text{lat}) * 10$ degrees of longitude are allowed to impact any particular state variable. This limitation is qualitatively identical to the cutoff radius employed by Houtekamer and Mitchell (1998). In later work, Houtekamer and Mitchell (2000) report that their use of a cutoff radius when using an EKF leads to discontinuities in the analysis. Here, this behavior was not observed, presumably because the EAF does not introduce the noise that can impact correlations in the EKF and because state variables that are adjacent on the grid are impacted by sets of observations that have a relatively large overlap. One could implement smoother algorithms for limiting the spatial range of impacts of an observation by multiplying the off-diagonal terms of the prior covariance matrix in (13) and (14) by an appropriate function of the distance between the observation and the state variable being updated; methods of this type are discussed in Hamill et al (2000).

Figure 8 shows time series of the truth, the ensemble mean and 10 of the ensemble members for a grid point near 45N 0E. Figure 9 shows the corresponding RMS error of the ensemble mean and the ensemble spread for the same variable. The RMS streamfunction error is consistently much less than the observational error standard deviation, even though only 250 observations are available. The truth generally stays inside the 10 ensemble members plotted in Fig. 8. The chi-square statistic for the bins over the 100 observation time interval from times 100 to 200 is 30.19, corresponding to a 93% chance that the truth was not picked from the same distribution as the ensemble. In general, for this assimilation, a sample of 100 observation times is enough to distinguish the truth from the ensemble at about the 90% confidence level. The ratio of R to its expected value is 1.026 indicating that in general this ensemble assimilation is somewhat too confident (too little spread).

Fig. 10 plots the error of the ensemble mean streamfunction field at assimilation time 200. All shaded areas have error magnitude less than the observational standard deviation. The largest errors are in the region between 60E and 160E where there are no observations. The areas of smallest error are concentrated in areas distant from and generally in regions upstream from the data void.

As noted in section 3, it is important to know something about the error growth of the model when the data assimilation is turned off in order to be able to judge the value of the assimilation method. For this barotropic model, the ensemble mean RMS error doubles in about 10 days.

The second case examined in this section uses the same T42 model (with the weak climatological forcing removed) to assimilate data from the NCEP operational analyses for the winter of 1991-92. The 'observations' are available once a day as T21 truncated spherical harmonics and are interpolated to the Gaussian grid points of the T42 model being used. This interpolation is regarded as the truth and observations are taken at each grid point by adding observational noise with a standard deviation of $1x10^6$ m^2s^{-1}. This is a particularly challenging problem for the EAF because the T42 model has enormous systematic errors at a lead time of 24 hours. The result is that the impact of the observations is large while the EAF is expected to work best when the impact of observations is relatively small (see section 2c).

In addition, the EAF as described to date assumes that the model being used has no systematic errors. That is obviously not the case here and a direct application of the filter method as described above does not work well. A simple modification of the filter to deal with model systematic error is to include an additional parameter that multiplies the prior covariance, Σ_p, only when it is used in (14) to compute the updated mean. Setting this factor to a value greater than 1 indicates that the prior estimate of the position of the mean should not be regarded as being as confident as the prior ensemble spread would indicate. In the assimilation shown here, this factor is set to 1.02. A covariance inflation factor must also continue to be used. Because error growth in the T42 barotropic model is much slower than that in the atmosphere, this factor is much larger here than in the perfect model cases and serves to correct the systematic slowness of uncertainty growth in the assimilating model. Covariance inflate is set to 1.45 here.

Figure 11 shows a time series of the truth, ensemble mean and 10 ensemble members from the T42 assimilation of NCEP data for stream function near 45N 0E, the same point shown in the perfect model results earlier in this section. The ensemble assimilation clearly tracks the observed data which have much higher amplitude and frequency temporal variability than is seen in the perfect model in Fig. 8. Although the truth frequently falls within the ten ensemble members, this variable has a chi-square statistic of 46.00 which gives 99% confidence that the truth is not drawn from the same distribution as the ensemble given 100 days of assimilation starting on 11 November, 1991. Given the low quality of the model, these results still seem to be reasonably good. Figure 12 plots the error of the ensemble mean on 19 February, 1992, a typical day. All shaded areas have ensemble mean error less than the observational error standard deviation with dark shaded regions having less than 25% of this error. These results give some encouragement that practical assimilation schemes for operational applications could be obtained if the EAF were applied with a more realistic forecast model and more frequent observations.

8. Conclusions and future development

The EAF can do viable data assimilation and prediction in models with state space dimension large compared to the ensemble size. It has an ability to assimilate observations with complex non-linear relations to the state variables and has extremely favorable computational scaling for large models. At least in low-order models, the EAF compares quite favorably to the four dimensional variational method, producing assimilations with smaller error and also providing information about the distribution of the assimilation. Unlike variational methods, the EAF does not require the use of linear tangent and adjoint model codes and so is straightforward to implement, at least mechanistically, in any prediction model. The EAF is similar in many ways to the EKF, but uses a different algorithm for updating the ensemble when observations become available. The EKF introduces noise by forming a random sample of the observational error distribution and this noise has an adverse impact on the quality of assimilations produced by the EKF.

It is possible that additional heuristic modifications to the EKF could make it more competitive with the EAF. Comparing the EAF to other methods in large models is impossible at present. Both of these points underscore the need to develop some sort of data assimilation testbed facility that allows experts to do fair comparisons of the many assimilation techniques that are under development.

The EAF can be extended to a number of other interesting problems. The version of the filter used here is currently being used in a study of adaptive observing systems (Berliner et al 1999, Palmer et al 1998). Just as the ensemble can provide estimates of the joint distribution of model state variables and observed variables, it can also provide estimates of joint distributions of the model state at earlier times with the state at the present time. Likewise, joint distributions of the state variables at different forecast times can be produced. These joint distributions can be used to examine the impact of observations at previous times, or during a forecast, on the state distribution at later times, allowing one to address questions about the potential value of additional observations (Bishop and Toth 1999). In a similar spirit, the ensemble filter provides a potentially powerful context for doing observing system simulation experiments (for instance Kuo et al 1998).

Another product of the filter assimilation is estimates of the covariances between state variables or state variables and observations (Ehrendorfer and Tribbia 1997). These estimates are similar to those that are required for simpler data assimilation schemes like optimal interpolation but also may be useful for theoretical understanding of the dynamics of the atmosphere (Bouttier 1993).

Despite the encouraging results presented here, there are a number of issues that must still be addressed before the EAF could be extended to application in operational atmospheric or oceanic assimilation. The most serious problem appears to be dealing with model uncertainty in a systematic way. In the work presented here, the covariance inflation factor has been used to prevent model prior estimates from becoming unrealistically confident. The current implementation works well in perfect model assimilations with homogeneous observations (observations of the same type distributed roughly uniformly in space), but begins to display some undesirable behavior with heterogeneous observations. In the barotropic model with a data void this was reflected as an inability to produce good RMS ratios in both the observed and data void areas. Reducing the covariance inflation factor when the spread for a state variable becomes large compared to the climatological standard deviation (not done in the results displayed here) solves this problem. Another example of this problem occurs when observations of both temperature and wind speed are available in primitive equation model assimilations. Clearly, a more theoretically grounded method for dealing with model uncertainty is needed. Nevertheless, the covariance inflation approach does have a number of desirable features that need to be incorporated in a more sophisticated approach. Operational atmospheric models tend to have a number of balances that constrain the relation between different state variables. If the problem of model uncertainty is dealt with in a naive fashion by just introducing some unstructured noise to the model, these balance requirements are ignored. As an example, in primitive equation applications, this results in excessive gravity wave noise in the assimilation (Anderson 1997). The covariance inflate approach maintains existing linear relations between state variables, and so produces far less gravity wave noise in primitive equation tests to date. The EKF introduces noise when computing the impact of observations on the prior state and this noise may also lead to increased gravity wave noise in assimilations.

Dealing with the more serious model errors that occur in assimilation of observed atmospheric data requires even more careful thought. Introducing an additional parameter that controls the confidence placed in prior estimates of the mean is able to deal with a number of model biases, but a more theoretically grounded approach would be desirable.

Ongoing work with the EAF is addressing these issues and gradually expanding the size and complexity of the assimilating models. Initial results with coarse resolution dry primitive equation models are to be extended to higher resolutions with moist physics. The filter is also scheduled to be implemented in the GFDL Modular Ocean Model for possible use in producing initial conditions for seasonal forecast integrations of coupled models.

Appendix A: Ensemble adjustment

This appendix describes a general implementation of the EAF; refer to the last paragraph of section 2 for details on how this method is applied in a computationally affordable fashion. Let $\{\mathbf{z}^p{}_i\}$, (i = 1, ..., N) be a sample of the prior distribution at a time when new observations become available with the subscript referring to each member of the sample (an N-member ensemble of state vectors). The prior sample mean and covariance are defined as $\bar{\mathbf{z}}^p$ and $\Sigma = \Sigma^p$. Assume that $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}^o$ and $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ are available at this time with $\mathbf{y}^o$ the observations vector, $\mathbf{R}$ the observational error covariance, and $\mathbf{H}$ a linear operator that produces the observations given a state vector.

Since Σ is symmetric, a singular value decomposition gives $\mathbf{D}^p = \mathbf{F}^T\Sigma\mathbf{F}$ where $\mathbf{D}^p$ is a diagonal matrix with the singular values, μ^p of Σ on the diagonal and $\mathbf{F}$ is a unitary matrix ($\mathbf{F}^*\mathbf{F}$ = I, $\mathbf{F}^{-1} = \mathbf{F}^*$, $(\mathbf{F}^*)^{-1} = \mathbf{F}$). Applying $\mathbf{F}^T$ and $\mathbf{F}$ in this fashion is a rotation of Σ to a reference frame in which the prior sample covariance is diagonal.

Next, one can apply a scaling in this rotated frame in order to make the prior sample covariance the identity. The matrix $\mathbf{G}^T\mathbf{F}^T\Sigma\mathbf{F}\mathbf{G}$, where $\mathbf{G}$ is a diagonal matrix with the square root of the singular values, μ^p, on the diagonal is the identity matrix, I.

Next, a singular value decomposition can be performed on the matrix $\mathbf{G}^{-1}\mathbf{F}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{F}(\mathbf{G}^{-1})^T$; this is a rotation to a reference frame in which the observational ‘covariance’ matrix, $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, is a diagonal matrix, $\mathbf{D} = \mathbf{U}^T\mathbf{G}^{-1}\mathbf{F}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{F}(\mathbf{G}^{-1})^T\mathbf{U}$ with the diagonal elements the singular vectors, μ. The prior covariance can also be moved to this reference frame, and it is still the identity since $\mathbf{U}$ is unitary, $\mathbf{I} = \mathbf{U}^T\mathbf{G}^T\mathbf{F}^T\Sigma\mathbf{F}\mathbf{G}\mathbf{U}$.

The updated covariance can be computed easily in this reference frame since the prior covariance inverse is just $\mathbf{I}$ and the observed ‘covariance’ is diagonal. The updated covariance can then be moved back to the original reference frame by unrotating, unscaling, and unrotating. (Note that $\mathbf{G}$ is symmetric).

More formally, the updated covariance can be evaluated as:

$$\Sigma^u = (\Sigma^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1} =$$

$$\mathbf{F}\mathbf{G}^{-1}\mathbf{U}\{\mathbf{U}^{-1}\mathbf{G}\mathbf{F}^{-1}(\Sigma^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}(\mathbf{F}^T)^{-1}\mathbf{G}^T(\mathbf{U}^T)^{-1}\}\mathbf{U}^T(\mathbf{G}^{-1})^T\mathbf{F}^T =$$

$$\mathbf{FG}^{-1}\mathbf{U}\{[\mathbf{U}^T(\mathbf{G}^T)^{-1}\mathbf{F}^T(\Sigma^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{FG}^{-1}\mathbf{U}]^{-1}\}\mathbf{U}^T(\mathbf{G}^{-1})^T\mathbf{F}^T =$$

$$\mathbf{FG}^{-1}\mathbf{U}\{[\mathbf{U}^T(\mathbf{G}^T)^{-1}\mathbf{F}^T\Sigma^{-1}\mathbf{FG}^{-1}\mathbf{U} + \mathbf{U}^T(\mathbf{G}^T)^{-1}\mathbf{F}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HFG}^{-1}\mathbf{U}]^{-1}\}\mathbf{U}^T(\mathbf{G}^{-1})^T\mathbf{F}^T =$$

$$\mathbf{FG}^{-1}\mathbf{U}\{[(\mathbf{U}^{-1}\mathbf{GF}^{-1}\Sigma(\mathbf{F}^T)^{-1}\mathbf{G}^T(\mathbf{U}^T)^{-1})^{-1} + \mathbf{U}^T(\mathbf{G}^T)^{-1}\mathbf{F}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HFG}^{-1}\mathbf{U}]^{-1}\}\mathbf{U}^T(\mathbf{G}^{-1})^T\mathbf{F}^T =$$

$$\mathbf{FG}^{-1}\mathbf{U}\{[(\mathbf{U}^T\mathbf{G}^T\mathbf{F}^T\Sigma\mathbf{FGU})^{-1} + \mathbf{U}^T(\mathbf{G}^T)^{-1}\mathbf{F}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HFG}^{-1}\mathbf{U}]^{-1}\}\mathbf{U}^T(\mathbf{G}^{-1})^T\mathbf{F}^T$$

The first inverse in parentheses is just I, and the second term in the square brackets is just diag[μ_1, μ_2, ...], so the term inside the curly brackets is diag[$1/(1 + \mu_1)$, $1/(1+\mu_2)$, ...].

The term inside the curly brackets can be rewritten as $\mathbf{B}^T\mathbf{U}^T\mathbf{G}^T\mathbf{F}^T\Sigma\mathbf{FGUB}$ where

$\boldsymbol{B} = diag[(1+\mu_1)^{-1/2}, (1+\mu_2)^{-1/2}, \ldots]\cdot$

Then, $\Sigma^u = \mathbf{A}\Sigma\mathbf{A}^T$ where $\mathbf{A}=\mathbf{FG}^{-1}\mathbf{UB}^T\mathbf{U}^T\mathbf{GF}^T$.

As noted above, being able to write an expression for Σ^u in this form enables an update of the prior sample, $\{\mathbf{z}^p_i\}$, to get an updated sample, $\{\mathbf{z}^u_i\}$ as:

$\mathbf{z}^u_i = \mathbf{A}^T(\mathbf{z}^p_i - \bar{\mathbf{z}}^p) + \bar{\mathbf{z}}^u$.

An understanding of this update process follows from the discussion above. After applying the rotation, scaling and rotation operators $\mathbf{U}^T$, $\mathbf{G}$, and $\mathbf{F}^T$, to the prior sample, it is in a space where the prior sample covariance is **I** and the observational 'covariance' is diagonal. One can then just 'shrink' the prior ensemble by the factor $1/(1+\mu_i)$ independently in each direction to get a new sample with the updated covariance in this frame. The rotations and scaling can then be inverted to get the final updated ensemble.

The mean of the updated distribution needs to be calculated to compute the $\mathbf{z}^u_i$. Once the updated sample covariance has been computed as outlined above, the mean is calculated easily as $\bar{\mathbf{z}}^u = \Sigma^u(\Sigma^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}^o)$. For computational efficiency, Σ^{-1} can be computed by transforming back from the rotated sample SVD space in which it is diagonal.

If the ensemble size, N, is not larger than the size of the state vectors then the sample covariance matrix is degenerate, i.e. there are directions in the state space in which the ensemble has no variance. Applying the SVD to such sample covariance matrices actually results in a set of $m < N$ non-zero singular values and N-m zeros on the diagonal of $\mathbf{D}^p$. All the computations can then be performed in the m-dimensional subspace spanned by the singular vectors corresponding to the m non-zero singular values. In addition, there may be some set of singular values that are very small but non-zero. If care is used, these directions can also be neglected in the computation for further savings.

Appendix B: The Lorenz 96 Model

The Lorenz 96 model is a variable size low-order dynamical system used by Lorenz (1996) and more recently by others including Lorenz and Emanuel(1998). The model has N state variables, $X_1, X_2, ..., X_N$ and is governed by the equation

$$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$$

where i = 1, ..., N with cyclic indices. The results shown are for parameters with a sensitive dependence on initial conditions: N = 40, F = 8.0, and a 4th-order Runge-Kutta timestep with dt=0.05 is applied as in Lorenz and Emanuel.

Appendix C: Non-divergent barotropic model

A spherical harmonic model of the non-divergent barotropic vorticity equation on the sphere is used with a transform method for nonlinear terms performed on a non-aliasing physical space grid with 128 longitudes and 64 Gaussian latitudes for a total of 8192 grid points. A timestep of 1800s is used with a 3rd order Adams-Bashforth timestep which is initialized with a single forward step followed by a single leapfrog step. A ∇^8 diffusion on the stream function is applied with a constant factor so that the smallest resolved wave is damped with an e-folding time of 2 days. When run in a perfect model setting, a forcing must be added to the model to induce interesting long-term variability. In this case, the zonal flow spherical harmonic components are relaxed towards the observed time mean zonal flow for the period November through March 1991-92, with an e-folding time of approximately 20 days.

Acknowledgements

The author would like to thank Tony Rosati, Matt Harrison, Steve Griffies, and Shaoqing Zhang for their comments on earlier versions of this manuscript. The two anonymous reviewers were unusually helpful in finding errors and improving the originally submitted draft. Conversations with Michael Tippett, Peter Houtekamer, Ron Errico, Tom Hamill, Jeff Whitaker, Lenny Smith and Chris Snyder led to modifications of many of the ideas underlying the filter. Finally, this work would never have proceeded without the insight and encouragement of Stephen Anderson.

Figure Captions

1. Schematic showing results of applying different filters to two variables X_1 and X_2 in different compute subsets. Panel (a) shows the prior distribution of an eight member ensemble in the X_1-X_2 plane and the solid curve is an idealized distribution for an observation of X_1. The results of applying a kernel resampling filter (b), a single Gaussian resampling filter (c), and an ensemble adjustment filter (d) are depicted in the same plane.

2. RMS error as a function of forecast lead time (lead time 0 is the error of the assimilation) for ensemble adjustment filters with a 10-member ensemble (lowest dashed curve) and a 20-member ensemble (lowest solid curve) and for four dimensional variational assimilations that use the model as a strong constraint to fit observations over a number of observing times. In generally descending order, the number of observation times used by the variational method is two (dotted), three (dash-dotted), four (dashed), five (solid), six (dotted), seven (dash-dotted), eight (dashed), ten (solid), twelve (dotted) and fifteen (dash-dotted).

3. Time series of ‘truth’ from long control run (solid grey), and ensemble mean (thick dashed) and ten of the twenty individual ensemble members (thin dashed) for variable X_1 of the Lorenz-96

model from assimilation times 115 through 140 using an ensemble adjustment filter (a) and an ensemble Kalman filter (b).

4. Time series of RMS error of ensemble mean from ensemble adjustment filter assimilation (dashed) and mean RMS difference between ensemble members and the ensemble mean (spread, solid) for variable X_1 of the Lorenz-96 model from assimilation times 100 through 200 of the same assimilation as in Fig. 3.

5. Rank probability histogram (Talagrand diagram) of the true solution for X_1 within the 20 member ensemble of the ensemble adjustment filter assimilation for assimilation times 101 to 200 corresponding to the time series in Fig. 4.

6. Time series of 'truth' from long control run (solid grey), and ensemble mean (thick dashed) and ten of the twenty individual ensemble members (thin dashed) for the model forcing variable, F, of the Lorenz-96 model from assimilation times 750 through 800 for an assimilation with nonlinear observations operator described in text; results from ensemble adjustment filter (a) and ensemble Kalman filter (b).

7. As in Fig. 6a for variable X_1 of the Lorenz-96 model from assimilation times 750 through 800 using an ensemble adjustment filter.

8. Time series of 'truth' from long control run (solid grey), ensemble mean from ensemble adjustment filter assimilation in global barotropic model (thick dashed), and ten of the twenty individual ensemble members (thin dashed) for streamfunction at 45N 0. Observations are available every 12 hours and consisted of 250 points placed randomly on the surface of the sphere excluding the longitude belt from 60E to 160E where there were no observations; the observational error standard deviation was $1 \times 10^6\ m^2 s^{-1}$.

9. Time series of RMS error of ensemble mean from ensemble adjustment filter assimilation (dashed) and mean RMS difference between ensemble members and the ensemble mean (spread, solid) for streamfunction at 45N 0 for the same assimilation as in Fig. 8.

10. Error of ensemble mean of assimilation at assimilation step 200 for the same assimilation as in Fig. 8. In addition to shading, contours are plotted with an interval of 1×10^6 for absolute value greater than 1×10^6.

11. Time series of ‘truth’ from NCEP analyses (solid grey), ensemble mean from ensemble adjustment filter assimilation (thick dashed), and ten of the twenty individual ensemble members (thin dashed) for streamfunction at 45N 0 from a T42 barotropic model. Observations are available at each model gridpoint once per day with observational error standard deviation of 1×10^6 $m^2 s^{-1}$.

12. Error of ensemble mean of assimilation at day 200 for the same assimilation as in Fig. 11. In addition to shading, contours are plotted with an interval of 1×10^6 for absolute value greater than 1×10^6.

References:

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518-1530.

Anderson, J. L., 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Wea. Rev.*, **125**, 2969-2983.

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.

Barker, T. W., 1991:The relationship between spread and forecast error in extended-range forecasts. *J. Climate*, **4**, 733-742.

Berliner, L. M., Z.-Q. Lu, and C. Snyder, 1999: Statistical design for adaptive weather observations. *J. Atmos. Sci.*, **56**, 2536-2552.

Bishop, C. H. and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748-1765.

Bouttier, F., 1993: The dynamics of error covariances in a barotropic model. *Tellus(A*), **45**, 408-423.

Brasseur, P., J. Ballabrera-Poy, and J. Verron, 1999: Assimilation of altimetric data in the mid-latitude oceans using the singular evolutive. *J. Marine Systems*, **22**, 269-294.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.

Courtier, P., J. Derber, R. Errico, J.-F. Louis, and T. Vukicevic: 1993: Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus(A)*. **45**, 342-357.

Derber, J. C.,1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2347-2446.

Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703-713.

Ehrendorfer, M. and J. J.Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci*., **54**, 286-313.

Evans, R. E., M. S. J. Harrison, R. J. Graham, and K. R. Mylne, 2000: Joint Medium-Range ensembles from the Met. Office and ECMWF systems. *Mon. Wea. Rev.*, **128**, 3104-3127.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res*., **99**, 10143-10162.

Evensen, G., and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.

Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187-1193.

Gordeau, L., J. Verron, T. Delcroix, A. J. Busalacchi, and R. Murtugudde, 2000: Assimilation of TOPEX/POSEIDON altimetric data in a primitive equation model of the tropical Pacific Ocean during the 1992-1996 El Nino-Southern Oscillation period. *J. Geophys. Res.*, **105**, 8473-8488.

Hamill, T. M., C. Synder, and R. J. Purser, 2000: Digital filtering of background error covariance estimates generated by a forecast ensemble. In press with Mon. Wea. Rev.

Hansen, J. A. and L. A. Smith, 2000: Probabilistic Noise Reduction. In press with Tellus.

Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: two transplant case-studies. *Quart. J. Roy. Met. Soc.*, **125A**, 2487-2515.

Houtekamer, P. L., L. Lefaivre, and J. Derome, 1995: *The RPN Ensemble Prediction System. Proc. ECMWF Seminar on Predictability, Vol. II*, 121-146.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

Houtekamer, P. L. and L. Lefaivre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev.*, **125**, 2416-2426.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

Houtekamer, P. L, and H. L. Mitchell, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.*, **102**, 27835-27860.

Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971-1981.

Krishnamurti, T. N., C. M. Kishtawal, T. E. Larow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multmodel superensemble. *Science*, **285**, 1548-1550.

Kuo, Y.-H., X. Zou, and W. Huang, 1998: The impact of Global Positioning System data on the prediction of an extratropical cyclone: an observing system simulation experiment. *Dynamics of Atmos. and Oceans*, **27**, 439-470.

Le Dimet, F.-X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, **38**, 97-110.

Lorenc, Andrew C., 1997: Development of an operational variational assimilation scheme. *J. Met. Soc. Japan*, 229-236.

Lorenz, E., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Vol. 1, ECMWF, Reading Berkshire, UK, 1-18.

Lorenz, E. N. and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399-414.

Miller, R. N., M. Ghil and F. Gauthiez, 1994a: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci*., **51**, 1037-1056.

Miller, R. N., E. D. Zaron, and A. F. Bennett, 1994b: Data assimilation in models with convective adjustment. *Mon. Wea. Rev.*, **122**, 2607-2613.

Mitchell, H. L. and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy Meteor. Soc.*, **122**, 73-120.

Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 463-493.

Murphy, J. M., 1990: Assessment of the practical utility of extended range ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **116**, 89-125.

Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633-653.

Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1-39.

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649-667.

Swanson, K., R. Vautard, and C. Pires, 1998: Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus(A)*, **50**, 369-390.

Tarantola, A., 1987: Inverse Problem Theory. Elsevier Science Publishers B. V., 613 pp.

Tziperman, E. and Z. Sirkes, 1997: Using the adjoint method with the ocean component of coupled ocean-atmosphere models. *J. Met. Soc. Japan*, **75**, 353-360.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319

Tracton, S., and E. Kalnay, 1993: Operational ensemble forecasting at NMC: Practical aspects. *Wea. Forecasting*, **8**, 379-398.

Van Leeuwen, P. J. and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.*, **124**, 2898-2912.

Van Leeuwen, P. J., 1999: Comment on ‘‘Data assimilation using an ensemble Kalman filter technique.’’ *Mon. Wea. Rev.*, **127**, 1374-1377.

Vukicevic, T. and R. M. Errico, 1993: Linearization and adjoint of parameterized moist diabatic processes. *Tellus(A)*, **45**, 493-510.

Ziehmann, C., 2000: Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus*, **52A**, 280-299.

Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, **125**, 2274-2292.